A Hybrid Discourse Relation Parser in CoNLL 2015

Sobha Lalitha Devi., Sindhuja Gopalan., Lakshmi S., Pattabhi RK Rao., Vijay Sundar Ram R., and Malarkodi C.S.

AU-KBC Research Centre

MIT Campus of Anna University

Chromepet, Chennai, India

sobha@au-kbc.org

Abstract

The work presented here describes our participation in CoNLL 2015 shared task in the closed track. Here we have used a hybrid approach, where Machine Learning (ML) technique and linguistic rules are used to identify the discourse relations. We have developed this system with a view that it consistently works across all domains and all types of text corpus. We have obtained encouraging results. The performance on blind test data and test data were similar.

1 Introduction

This paper describes our system, used in CoNLL-2015 shared task "Shallow Discourse Parsing". The goal of this task is to parse a piece of text into a set of discourse relations that exist between two adjacent or non-adjacent discourse units. Discourse relations are the coherence relations between two sentences that can be realized explicitly or implicitly in a text. Discourse connectives play a role in signaling the relations in a discourse. They connect two discourse units, which may be a sentence, clause or multiple sentences. These units are called arguments. Hence a discourse relation includes the connective and its arguments. The relations can be intra sentential or inter sentential i.e. it can occur within a sentence or across sentences.

Penn Discourse Tree Bank (PDTB) is used as the shared task data set for training and development. For the testing the shared task organizers have provided a blind set data, which is not from PDTB. PDTB is a richly annotated resource for discourse relations and their arguments. To develop PDTB, 1 million words Wall Street Journal is used as a corpus. It is annotated with five types of relations, Explicit, Implicit, EntRel, AltLex and NoRel. Discourse relations in PDTB are broadly classified into two types based on how the relations are realized in the text. When the relation is realized explicitly by a lexical item that belongs to syntactically well defined classes, those connectives are classified as "Explicit connectives". If a relation exists between adjacent sentences in the absence of explicit markers, "Implicit relation" can be inferred.

The main objective of the work presented here is to develop a system for identifying different types of discourse relations automatically. We have followed a hybrid approach, where we first use Machine Learning (ML) technique to identify the discourse relations and then enhance the results using a rule based approach. In the following sections, we give a detailed description of our system.

2 Explicit Relation Identification

Discourse relation is realized by Explicit connectives between two discourse units. The discourse units can be a clause, sentence or multiple sentences. The units they connect are referred as argument 1 and argument 2. Explicit connectives mainly belong to three syntactic classes, which include Subordinating conjunction, Coordinating conjunction and Discourse adverbials. PDTB provides sense classification for Explicit, Implicit and AltLex relations. Discourse connectives are broadly classified into four classes based on science.

a) Expansion b) Contingency c) Temporal, d). Comparison.

In order to refine the sense classification further, each class is defined with further types and subtypes. In this paper, we present a hybrid system for automatic identification of connectives and their arguments from parse text, developed using graph based machine learning technique CRFs and linguistic rules.

CRFs is a finite state model with un-normalized transition probability. It solves label bias problem efficiently. It has a single exponential model for joint probability of the entire sequence of labels when an observation sequence is given (Lafferty et al, 2001). The true power of graphical models lies in their ability to model many variables that are independent of each other (Sutton et al, 2011). For our work we have used the CRF++, which is a simple and customizable tool (Kudo, 2005).

The identification of explicit relations includes two subtasks, 1. Connective identification and classification 2. Argument identification and extraction. The discourse relations occur as inter-sentential or intra-sentential in a text. First, our system identifies whether a connective exist as discourse connective in the context. Consider the below example,

Example [1]

Morgan Stanley and Kidder Peabody, the two biggest program trading firms, staunchly defend their strategies.

In Example [1], the lexical item "and" is not a discourse connective but acts as conjunction joining two nouns Morgan Stanley and Kidder Peabody. Hence it is important to identify whether the connective acts as discourse connective or not in a context.

After identifying the discourse connective, the system predicts its sense. One connective can have multiple senses.

Example [2]

Several big securities firms backed off from program trading a few months after the 1987 crash . But most of them, led by Morgan Stanley & Co., moved back in earlier this year. ("But" Sense: Comparison.Contrast)

Example [3]

Just the thing for the Vivaldi-at-brunch set, the yuppie audience that has embraced New Age as its very own easy listening. But you can't dismiss Mr. Stoltzman's music or his motives as merely commercial and lightweight. ("But" Sense: Comparison.Concession)

In above Examples [2] & [3], "But" acts as an inter sentential connective. Although "But" in the above examples is syntactically similar, it has a different sense.

In these examples "But" acts as comparitive connective, but vary in its type. In the CoNLL version of PDTB data "but" with the sense

"Comparison.Contrast" occurred in 70.48% cases. In some cases, the sense for a connective may vary even at class level.

After identifying and predicting the sense of a connective, the span of arguments they connect needs to be identified. It is not necessary that the relation should occur between adjacent sentences. It may span across sentences. However, PDTB follows a minimality principle for annotating the arguments. The minimal information required to complete the interpretation of the arguments is annotated.

2.1 System description

Motivated by the work of Lin et al (2009), we have designed our system as a pipeline, where the relations are identified in sequential order. First, the system identifies and predicts the discourse connectives and their sense. Then, using the identified connectives argument 1 and argument 2 spans are identified and extracted. Then, the system examines all sentence pairs. The pair that is not identified in explicit relation is then classified into Implicit, Entrel or Altlex relation by the system.

2.2 System description Connective Identification and Sense Prediction

In the task of connective identification, the system is first trained to identify the connectives syntactically i.e. to identify whether the connective functions as a discourse connective or not. Then, the connectives are classified based on its sense. We have extracted the word and other syntactic features such as POS, chunk and Clausal information from PDTB parse text. In the task of identifying the discourse connectives, the system is trained using lexico-syntactic features like Word, Parts-of-speech (POS), Chunk, Combination of word, POS and chunk and Clause in a window size of 3. The lexicon itself acts as a good feature to identify the discourse connectives. POS, chunk and clausal information help in disambiguating the connectives.

Example [4]

after IN B-PP Temporal.Asynchronous.Succession interviewing VBG B-VP o

Generally, "after" exists as connective and also as preposition or adverbs in a corpus. But when "after" is followed by a gerund, it acts as discourse connective. The POS for a gerund is "VBG" and hence plays an important role in dis-

course connective identification. The clausal information also helps in identifying a lexical unit as discourse connective because when a discourse connective exists in a sentence, then it will be mostly succeeded or preceded by a clause. In addition to these features, we have used dictionary inside the CRFs. We have developed the dictionary based on connectives that are not ambiguous. After identifying the connectives, we analysed the errors generated by the system. We found the system has tagged the connectives that are not discourse connectives. Hence it resulted in false positives.

Example[5]
Our offer is to buy any and all shares tentered at $18 a share.

In the above example "and" is not a discourse connective, but the system tagged wrongly discourse connective.

Example [6]
A spokesman for Dow Jones said he hadn't seen the group's filing, but added, ``obviously Dow Jones disagrees with their conclusions.

In the above example the connective "but" was not identified by the system. Hence, we used post processing rules to improve the connective identification.

Once the discourse connectives are identified, the system predicts the sense of the connectives. Using the above mentioned lexico-syntactic features and connectives, we developed individual models for each type of sense. In the case of sense identification, connective itself is a good feature, as only few connectives are ambiguous. To solve the ambiguity in the case of sense classification, the preceding and succeeding POS and words were useful to some extent. Using these models, senses of connectives are identified separately. Then we merged the output based on the confidence scores.

Error analysis on sense classification showed that the sense is wrongly predicted by the system. Consider the below example [7], where "until" is predicted as "Contingency.Condition" by the system, but the sense of the connective "until" is "Temporal.Asynchronous.Precedence"

Example [7]
He's an ex-hurler who's one of the leading gurus of the fashionable delivery, which looks like a fastball until it dives beneath the lunging bat.

Heuristic based post processing rules were used to correct and improve the sense prediction.

2.3 Argument identification

In the next phase, the system is trained to identify the arguments and their text spans. We have followed the method used by Menaka et al (2011) for identification of causal relations from Tamil data. In their work, instead of identifying the whole argument, the boundaries of the arguments were identified. Similarly, we created individual model for each boundary, i.e. for Argument 1 start, Argument 1 end, Argument 2 start and Argument 2 end. The connective tagged input is given for argument extraction. For argument identification we have developed separate models for inter and intra sentential relation. Each connective is processed separately and is given as input to inter sentential and intra sentential models. We have used the following features for identifying the argument boundaries.

    a. Word , POS, Chunk
    b. Combination of word, POS, Chunk
    c. Clausal boundaries
    d. Sentence boundaries
    e. Connective.

We have used connectives as features, as the argument 2 start and argument 1 end are syntactically associated with the connective in most of the cases. Hence, when the connective is identified, the position of Argument 2 start and Argument 1 end boundary can be located. In most of the cases the Argument 1 start is present at the initial position of a sentence or clause and Argument 2 end at the final position of a sentence or clause. In the case of inter sentential relation, the previous sentence to the connective acts as Argument 1. Here, the sentence final position acts as Argument 1 end. Therefore, sentence and clausal boundaries are used as features for argument identification in our work. After identifying the argument boundaries separately, we merged the output from four language models. In order to improve the system's performance for argument extraction further, we used linguistic and heuristic rules. In the following paragraph, we describe some of the linguistic and heuristic rules.

Rules Description

Example [8]
At Shearson Lehman, executives created potential new commercials Friday night and throughout the weekend, then had to regroup yesterday afternoon.

In the above example Argument 2 end was not marked by the system. In such case we used heuristic rule to identify the Argument 2 end boundary.

Example [9]

The agency has already spent roughly $ 19 biollion selling 34 insolvent S&Ls, and it is likely to sell or merge 600 by the time the bailout concludes.

The above Example [9] is a simple discourse relation that exists in the corpus. Using simple linguistic rules, such relations can be identified. In this case, when punctuation mark "," (comma) is followed by a connective; the span above comma is marked as Argument 1 and the span below connective is marked as Argument 2.

3   Non-Explicit Relation Identification

In the task of Non-Explicit relation identification, we identify the sentences which can possibly have implicit relations, AltLex and EntRel relations. And then the sense of the Implicit connective and AltLex is identified. The identification of implicit relation between a pair of sentences is done using a machine learning technique, CRFs. From the input data we look for sentences without Explicit connectives and form pair of sentences by considering its previous sentence. Features extracted from this pair of sentences are given to the CRFs engine to identify the presence of implicit relation. We use the following features:

i.   **Presence of common words:** The count of commonly occurring words in the argument 1 and argument 2 is taken. Here we remove the stop words.

ii.   **Difference in the polarity:** The average polarity of each sentence is calculated. First each word is marked with its polarity score as obtained from the MPQA polarity lexicon provided by the task organizers. The average score of each sentence in the pair is calculated by aggregating the individual word scores. If the polarities are same in both sentences, then the feature is given the value of 0:0, if sentence 1 has positive score and sentence two has negative score, then feature is given a value of 1:-1, else viceversa.

iii.   Commonality of the words in the initial and terminal positions of the sentences

iv.   Presence of common brown cluster IDs

v.   Presence of common bigrams and trigrams

The output obtained from the machine learning engine is given the secondary engine. In the secondary engine, we check the coreference between the pair of sentence using anaphora resolution system. Those pair of sentences which have common coreference mentions we consider this pair of sentences to have implicit relations.

We have used an in-house developed anaphora resolution system (Sobha, 2011), which uses salience measure based approach.

Thus we identify the sentence pairs which have the implicit relations in them. The next task is to identify the sense of the Implicit connective between this pair of sentences. For the purpose of identifying the sense (i.e., sense classification), we first identify or learn common patterns from the Implicit and Explicit sense annotated training data. And these patterns are given as features to the CRFs machine learning, which would finally mark the sense of the implicit relation. In the previous reported works we observe that most of the sense classification was restricted to four top level senses i.e., Expansion, Contingency, Comparison and Temporal, whereas in our present work we need to identify the senses to finer granularity levels; such as "Expansion.Alternative.Chosen alternative", "Contingency.Cause.Result". Thus, this leads to the 14 different senses.

The common patterns in the Explicit and Implicit training data are learned based on the two factors Polarity scores and the verb clusters obtained from the VerbNet. The patterns are formed by considering two factors from argument 1 and argument 2 and a tuple is formed. This tuple consists <Verb_class of argument 1, Polarity of argument 1, Verb_class of argument 2, Polarity of argument 2, Associated sense> The number of common patterns learned from the Explicit and Implicit annotated training data is observed to be 535 unique patterns. And it has been observed in the data that also majority of these patterns is majorly associated with the senses "Expansion.Conjunction" (48.03%), "Expansion.Restatement" (17.19%), "Comparison.Contrast" (15.14%).

When we used these learned patterns on the development data to identify the similarities of the patterns, we obtained only a similarity of 35% of the patterns. This showed that sense classification in implicit relations is very much subjective

53

and depends on the semantics of the arguments argument 1 and argument 2. But in this work we have restricted ourselves with syntactic features and patterns as described earlier for developing a CRFs machine learning system for sense classification. The other syntactic features used are Part-of-speech (POS tags), First-Last-First three words of the arguments, bigrams and trigrams of POS tags, count of common brown cluster IDs. The features of First-last-first three words, count of common brown clusters, polarity score are used as described in (Pitler et al., 2009; Lin et al., 2009; Louis et al., 2010; Zhou et al., 2010). For the pair of sentences for which the sense has been identified, the first sentence is tagged as Argument 1 and the second sentence is tagged as Argument 2.

4 Results

In the table 1, we show the results obtained for our system for Explicit and Non-Explicit relations and overall.

We can observe from the results identifying implicit relation has been a harder task.

In the argument identification sub task we observe that identification of argument boundaries which are farther from the connective had been tough. Since the PDTB follows the principle of minimality, identifying the minimal span by system was not possible in 30% of the cases. This was due to the fact that we were only using syntactic features for learning. Since argument 2 was syntactically bound to the connective in most of the cases, the system could learn the argument 2 span better than the argument 1 span.

The system failed to identify correct Argument 1 span in cases where coordinating conjunction is the connective and Argument 1 span crosses more than two clauses or sentences. Especially for the connectives "and" and "or" identifying Argument 1 span has been ambiguous.

In this work we have restricted or assumed that in inter-sentential connective Argument 1 and Argument 2 spans are within the current and previous sentences and does not cross $(n-1)^{th}$ sentence. Though in reality, there are more than 5% cases which have an argument span of more than n and $(n-1)^{th}$ sentence. This assumption was made because more than 90% of the connectives which are inter sentential have a span of only two sentences and was also computationally simple.

5 Conclusion

This paper describes our participation in CoNLL 2015 shared task of Shallow discourse parsing. We have developed an automatic system which identifies different discourse relations along with their senses. Our main objective was to develop a system which works consistently across any given corpus or text. And we find that our system has performed consistently with same performance metrics for both PDTB test section and blind test set provided by the task organizers. We have obtained an overall F1 score for the discourse parser as 0.1502, precision of 0.159 and recall of 0.1423. The scores are encouraging.

Details	Blind Test Data			PDTB Test Data			Development Data		
	F1	Precision	Recall	F1	Precision	Recall	F1	Precision	Recall
All Arg 1 Argument 2 extraction	0.3317	0.3512	0.3143	0.3126	0.3176	0.3079	0.439	0.4407	0.4373
All Argument 1 extraction	0.3998	0.4233	0.3788	0.3807	0.3867	0.3749	0.5173	0.5193	0.5153
All Argument 2 extraction	0.5447	0.5767	0.5161	0.4624	0.4697	0.4554	0.59	0.5923	0.5877
All Explicit connective	0.8449	0.9232	0.7788	0.8644	0.9436	0.7974	0.928	0.9804	0.8809
All Sense	0.1232	0.1357	0.1253	0.132	0.2579	0.1284	0.213	0.4087	0.203
All Parser	**0.1502**	**0.159**	**0.1423**	**0.1461**	**0.1484**	**0.1439**	**0.2635**	**0.2646**	**0.2625**
CoNLL Baseline – All Parser	**0.0386**	**0.0376**	**0.0397**	**0.0306**	**0.0285**	**0.033**	-	-	-
Explicit only Arg 1 Argument 2 extraction	0.3473	0.3795	0.3201	0.3077	0.3359	0.2839	0.5469	0.5777	0.5191
Explicit only Argument 1 extraction	0.4449	0.4861	0.4101	0.3664	0.4	0.338	0.629	0.6645	0.5971
Explicit only Argument 2 extraction	0.642	0.7015	0.5917	0.4968	0.5423	0.4583	0.7591	0.802	0.7206
Explicit only Explicit connective	0.8449	0.9232	0.7788	0.8644	0.9436	0.7974	0.928	0.9804	0.8809
Explicit only Sense	0.2175	0.2639	0.2028	0.1924	0.347	0.1779	0.3745	0.5425	0.3494
Explicit only Parser	**0.2673**	**0.2921**	**0.2464**	**0.2678**	**0.2923**	**0.247**	**0.4911**	**0.5188**	**0.4662**
CoNLL Baseline – Explicit only Parser	**0.0**	**1.0**	**0.0**	**0.0**	**1.0**	**0.0**	**-**	**-**	**-**
Non-Explicit only Arg 1 Argument 2 extraction	0.3191	0.3295	0.3093	0.3166	0.3045	0.3297	0.3503	0.3378	0.3638
Non-Explicit only Argument 1 extraction	0.357	0.3687	0.3461	0.3828	0.3682	0.3986	0.4089	0.3943	0.4246
Non-Explicit only Argument 2	0.466	0.4812	0.4518	0.4329	0.4164	0.4508	0.4683	0.4349	0.4683

extraction									
Non-Explicit only Sense	0.0393	0.2081	0.043	0.0301	0.1015	0.0334	0.0643	0.2122	0.0647
Non-Explicit only Parser	**0.0553**	**0.0571**	**0.0536**	**0.0482**	**0.0464**	**0.0502**	**0.0764**	**0.0737**	**0.0794**
CoNLL Baseline – Non-Explicit only Parser	**0.0498**	**0.0376**	**0.0735**	**0.0393**	**0.0285**	**0.063**	-	-	-

Table 1. System Results for Blind Test Data, PDTB Test Data and Development data – This shows the results for all three different types of text corpora. Also, this shows the results for Explicit and Non Explicit relations separately.

Reference

Lafferty, J., McCallum, A., Pereira, F. 2001. *Conditional Random Fields:Probabilistic Models for Segmenting and Labeling Sequence Data*. In International Conference on Machine Learning., pages.1-8.

Ziheng Lin, Min-Yen Kan, and Hwee Tou Ng. 2009. *Recognizing implicit discourse relations in the Penn discourse treebank*. In EMNLP, pages 343–351.

Annie Louis, Aravind K. Joshi, Rashmi Prasad, and Ani Nenkova. 2010. *Using entity features to classify implicit discourse relations*. In SIGDIAL Conference, pages 59–62.

Menaka S., Pattabhi RK Rao., Sobha Lalitha Devi. 2011.*Automatic Identification of Cause-Effect Relations in Tamil Using CRFs*, In A. Gelbukh (ed), Springer LNCS Vol. 6608/2011 pp 316-327

Emily Pitler, Mridhula Raghupathy, Hena Mehta, Ani Nenkovak, Alan Lee, and Aravind K. Joshi. 2008. *Easily identifiable discourse relations*. In COLING (Posters), pages 87–90

Rashmi Prasad, Nikhil Dinesh, Alan Lee, Eleni Miltsakaki, Livio Robaldo, Aravind Joshi, Bonnie Webber. 2008. *The Penn Discourse TreeBank 2.0.* In Proceedings of LREC, 2008.

Sobha Lalitha Devi, Vijay Sundar Ram., Pattabhi RK Rao. 2011. *Resolution of Pronominal Anaphors using Linear and Tree CRFs*. In Proceedings of 8th DAARC, Faro, Portugal.

Charles Sutton and Andrew McCallum. 2011. *An Introduction to Conditional Random Fields*. Foundations and Trends in Machine Learning, Vol. 4(4), pages 267–373.

Ben Wellner, Lisa Ferro, Warren R. Greiff, and Lynette Hirschman. 2006. *Reading comprehension tests for computer-based understanding evaluation.* Natural Language Engineering, 12(4):305–334.

Zhi-Min Zhou, Yu Xu, Zheng-Yu Niu, Man Lan, Jian Su, and Chew Lim Tan. 2010. *Predicting discourse connectives for implicit discourse relation recognition*. In COLING (Posters), pages 1507–1514.

Kudo T. 2005. CRF++, an open source toolkit for CRF, http://crfpp.sourceforge.net